TESTE QUI-QUADRADO

1. Introdução

Geralmente, nem sempre se cumprem os requisitos exigidos para a utilização dos testes paramétricos (teste t, teste F, etc.), e por isso necessitamos outra série de provas para as quais não se exige que a distribuição populacional tenha uma forma determinada, nem que seus parâmetros cumpram certas condições numéricas como se exige nos testes paramétricos. Estas provas se conhecem com a denominação de *testes não paramétricos* e podem ser aplicadas a escalas nominais, ordinais ou de intervalo.

Dentro dos testes não paramétricos, um dos mais utilizados é o teste Qui-quadrado (χ^2), já que nos permite comparar estados nominais por meio de frequências observadas das que seriam de esperar, se fossem verdadeiras as condições hipotéticas do que se compara.

2. Teste qui-quadrado em tabelas de contingência

A classificação de observações (em geral, de variáveis qualitativas) de acordo com dois critérios é referida como tabela de contingência.

Se um critério envolve r categorias (linhas) e as outras c categorias (colunas), a tabela é referida como tabela $r \times c$. As vezes, cada uma das duas variáveis qualitativas pode dividir-se em só duas categorias ou níveis. Quando isto acontece, o resultado é uma tabela de contingência que consta de duas linhas e duas colunas. Tal tabela é conhecida por tabela de 2×2.

As Tabelas de contingência são construídas com o propósito de:

(1) Testar a relação de dependência (associação) entre duas variáveis (Teste de independência),
(2) Verificar se a distribuição de uma variável categórica é a mesma em diferentes populações (Teste de homogeneidade)

2.1. Teste de independência

O teste de independência é baseado no esquema amostral, no qual uma única amostra aleatória de tamanho n é classificada com relação a duas características simultaneamente;

Exemplo: Um grupo de investigadores publicou os resultados de um ensaio clínico do uso de certa medicação para a cura de uma certa doença. Pretende-se saber se existe associação entre o uso da medicação e cura. Seja $\alpha = 5\%$.

Medicação	Situação		Total
	Curado	Não curado	
Sim	90	10	100
Não	60	40	100
Total	150	50	200

Os passos a seguir para testar a associação entre a medicação e a cura das duas variáveis são:

1. **Estabelecer as hipóteses**
   $\boldsymbol{H_0}$: A cura é independente da medicação, ou seja,
   não há associação entre a cura e a medicação

   $\boldsymbol{H_1}$: A cura é dependente da medicação, ou seja,
   há associação entre a medicação e a cura

2. **Escolher o nível de significância:** Neste exemplo α = 0,05

3. **Calcular as frequências esperadas**

$$\textit{Curado com medicação}: \quad a = \frac{100 \times 150}{200} = 75$$

$$\textit{Não curado com medicação}: \quad b = \frac{100 \times 50}{200} = 25$$

$$\textit{Curado sem medicação}: \quad c = \frac{100 \times 150}{200} = 75$$

$$\textit{Não curado sem medicação}: \quad d = \frac{100 \times 50}{200} = 25$$

Assim, se as duas variáveis fossem independentes, nós esperaríamos as frequências que se apresentam entre parênteses na tabela:

Medicação	Situação		Total
	Curado	Não curado	
Sim	90 (75)	10 (25)	100
Não	60 (75)	40 (25)	100
Total	150	50	200

4. **Calcular o valor da estatística de prova**

$$\chi^2 = \sum \frac{(O - E)^2}{E}$$ onde ***O*** é o valor observado e ***E***, é o número esperado

Assim,

$$\chi^2 = \frac{(90 - 75)^2}{75} + \frac{(10 - 25)^2}{25} + \frac{(60 - 75)^2}{75} + \frac{(40 - 25)^2}{25} = 24$$

5. **Determinar o valor crítico**
   O valor crístico é obtido a partir das tabelas da distribuição de qui-quadrado para um dado valor α e n graus de liberdade (gl). Os graus de liberdade são obtidos a partir da fórmula: $gl = (r-1) \times (c-1)$.

   Para o caso deste exemplo, $gl = (2-1) \times (2-1) = 1$

   Assim: $\chi^2_{(1-\alpha,\ gl)} = \chi^2_{(0{,}95,\ 1)} = 3{,}841$

6. **Estabelecer a regra de decisão**

   Se $\chi^2_{calculado} > \chi^2_{crítico} \Rightarrow$ rejeita-se H_0
   Se $\chi^2_{calculado} < \chi^2_{crítico} \Rightarrow ac$eita-se H_0

7. **Concluir:**
   Dado que $\chi^2_{calc} > \chi^2_{crítico}$, rejeita-se a hipótese nula e conclui-se que existe uma associação estatisticamente significativa entre o uso da medicação e a cura, ao nível de 5%.

2.2. Teste de homogeneidade

O teste de homogeneidade é usado quando se deseja verificar se a distribuição de uma variável categórica é a mesma em diferentes populações, isto é, que as várias colunas (ou linhas) têm a mesma proporção de indivíduos nas várias categorias de uma característica, se os totais das linhas (ou colunas) são especificados antecipadamente.

Consideremos os dados do exemplo que se segue, para testar a igualdade das proporções de acasalamentos fecundos (e não fecundos) em três raças bovinas:

Raça	Tipo de acasalamento		Total
	Fecundo	Infecundo	
Charolesa	110	50	160
Gir	70	10	80
Nelore	30	10	40
Total	210	70	280

Vejamos os passos a seguir:

1. **Estabelecer as hipóteses**

A hipótese nula de homogeneidade que a proporção de cada tipo de acasalamento é a mesma para todas as raças, pode ser formalmente estabelecida como:

H_0: A proporção de acasalamentos fecundos é a mesma na três raças,
ou seja, $p_{Ch} = p_{Gir} = p_{Ne}$

H_1: As proporções não são todas iguais

2. **Definir o nível de significância:** Por exemplo $\alpha = 0{,}05$

3. **Calcular as frequências esperadas**

$$Charolesa\ fecundo = \frac{160 \times 210}{280} = 120 \qquad Charolesa\ infecundo = \frac{160 \times 70}{280} = 40$$

$$Gir\ fecundo = \frac{80 \times 210}{280} = 60 \qquad Gir\ infecundo: = \frac{80 \times 70}{280} = 20$$

$$Nelore\ fecundo: = \frac{40 \times 210}{280} = 30 \qquad Nelore\ infecundo: = \frac{40 \times 70}{280} = 10$$

Todos os valores calculados das frequências estão entre parênteses na tabela.

Raça	Tipo de acasalamento		Total
	Fecundo	Infecundo	
Charolesa	110 (120)	50 (40)	160
Gir	70 (60)	10 (20)	80
Nelore	30 (30)	10 (10)	40
Total	210	70	280

4. **Calcular o valor da estatística**

$$\chi^2 = \sum \frac{(O-E)^2}{E} = \frac{(110-120)^2}{120} + \cdots + \frac{(10-20)^2}{20} + \frac{(10-10)^2}{10} = 10$$

5. **Determinar o valor crítico**

   Para $gl = (3-1) \times (2-1) = 2$
   Temos: $\chi^2_{(1-\alpha,\ gl)} = \chi^2_{(0{,}95,\ 2)} = 5{,}991$

6. **Estabelecer a regra de decisão**

   Rejeitar H_0 se $\chi^2_{calculado} > \chi^2_{crítico} = 5{,}991$

7. **Concluir:**
   Como $\chi^2_{calculado} > \chi^2_{crítico}$, rejeita-se H_0 e conclui-se que as fecundidades das raças não são todas estatisticamente iguais, ao nível de 5%.

   Como H_0 foi rejeitada, deve-se continuar a investigação, comparando-se as raças duas a duas, para verificar quem difere de quem em termos do critério analisado.